

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

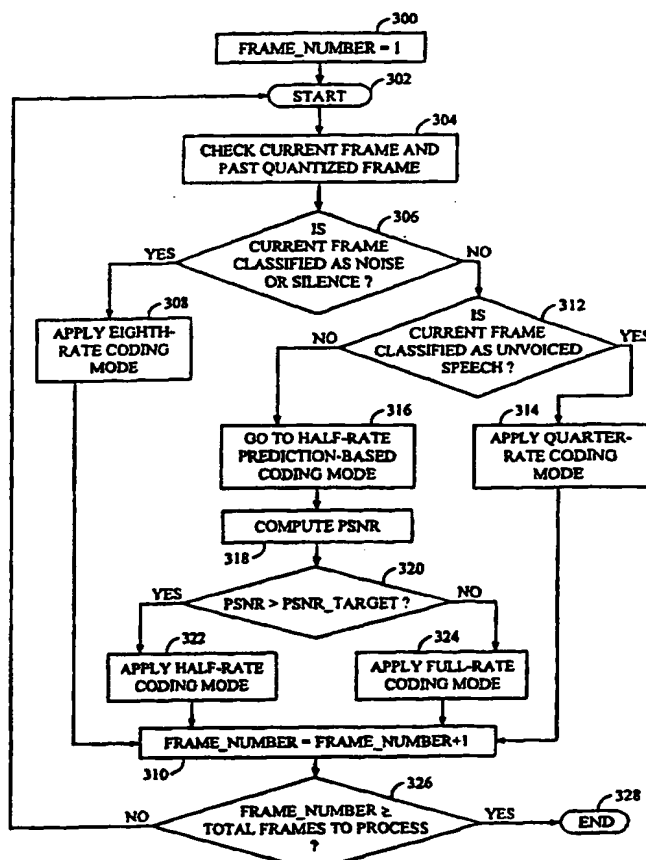
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G10L 19/14		A1	(11) International Publication Number: WO 00/30075
			(43) International Publication Date: 25 May 2000 (25.05.00)
(21) International Application Number: PCT/US99/26850 (22) International Filing Date: 12 November 1999 (12.11.99) (30) Priority Data: 09/191,643 13 November 1998 (13.11.98) US (71) Applicant: QUALCOMM INCORPORATED [US/US]; 5775 Morehouse Drive, San Diego, CA 92121-1714 (US). (72) Inventors: DAS, Amitava; 11738 Ramsdell Court, San Diego, CA 92131 (US). MANJUNATH, Sharath; 7104 Schilling Avenue #5, San Diego, CA 92126 (US). DEJACO, Andrew, P.; 10424 Flanders Cove, San Diego, CA 92126 (US). (74) Agents: MILLER, Russell, B. et al.; QUALCOMM Incorporated, 5775 Morehouse Drive, San Diego, CA 92121-1714 (US).			(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

(54) Title: CLOSED-LOOP VARIABLE-RATE MULTIMODE PREDICTIVE SPEECH CODER

(57) Abstract

A closed-loop, multimode, predictive speech coder includes a codec (100, 200) configured to operate in any of several coding modes, and a closed-loop mode decision module configured to apply a lowest-bit-rate coding mode to an input speech frame. A performance measure of the codec is obtained and compared with a threshold value. If the performance measure does not exceed the threshold value, the lowest-bit-rate coding mode is rejected in favor of a coding mode with a higher bit rate. The process can be continued until the coding performance is satisfactory. A higher-bit-rate, direct coding mode may be applied after a lower-bit-rate, prediction-based coding mode has failed to perform satisfactorily.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

CLOSED-LOOP VARIABLE-RATE MULTIMODE PREDICTIVE SPEECH CODER

5

BACKGROUND OF THE INVENTION

I. Field of the Invention

10 The present invention pertains generally to the field of speech processing, and more specifically to closed-loop, variable-rate, multimode, predictive coding of speech.

II. Background of Invention

15

Transmission of voice by digital techniques has become widespread, particularly in long distance and digital radio telephone applications. This, in turn, has created interest in determining the least amount of information that can be sent over a channel while maintaining the perceived quality of the reconstructed
20 speech. If speech is transmitted by simply sampling and digitizing, a data rate on the order of sixty-four kilobits per second (kbps) is required to achieve a speech quality of conventional analog telephone. However, through the use of speech analysis, followed by the appropriate coding, transmission, and resynthesis at the receiver, a significant reduction in the data rate can be achieved.

25

Devices that employ techniques to compress speech by extracting parameters that relate to a model of human speech generation are called speech coders. A speech coder divides the incoming speech signal into blocks of time, or analysis frames. Speech coders typically comprise an encoder and a decoder, or a codec. The encoder analyzes the incoming speech frame to extract certain relevant
30 parameters, and then quantizes the parameters into binary representation, i.e., to a set of bits or a binary data packet. The data packets are transmitted over the communication channel to a receiver and a decoder. The decoder processes the data packets, unquantizes them to produce the parameters, and then resynthesizes the speech frames using the unquantized parameters.

35

The function of the speech coder is to compress the digitized speech signal into a low-bit-rate signal by removing all of the natural redundancies inherent in

speech. The digital compression is achieved by representing the input speech frame with a set of parameters and employing quantization to represent the parameters with a set of bits. If the input speech frame has a number of bits N_i and the data packet produced by the speech coder has a number of bits N_o , the compression factor achieved by the speech coder is $C_r = N_i/N_o$. The challenge is to retain high voice quality of the decoded speech while achieving the target compression factor. The performance of a speech coder depends on (1) how well the speech model, or the combination of the analysis and synthesis process described above, performs, and (2) how well the parameter quantization process is performed at the target bit rate of N_o bits per frame. The goal of the speech model is thus to capture the essence of the speech signal, or the target voice quality, with a small set of parameters for each frame.

One effective technique to encode speech efficiently at low bit rate is multimode coding. A multimode coder applies different modes, or encoding-decoding algorithms, to different types of input speech frames. Each mode, or encoding-decoding process, is customized to represent a certain type of speech segment (i.e., voiced, unvoiced, or background noise) in the most efficient manner. An external mode decision mechanism examines the input speech frame and make a decision regarding which mode to apply to the frame. Typically, the mode decision is done in an open-loop fashion by extracting a number of parameters out of the input frame and evaluating them to make a decision as to which mode to apply. Thus, the mode decision is made without knowing in advance the exact condition of the output speech, i.e., how similar the output speech will be to the input speech in terms of voice-quality or any other performance measure. An exemplary open-loop mode decision for a speech codec is described in U.S. Patent No. 5,414,796, which is assigned to the assignee of the present invention and fully incorporated herein by reference.

Multimode coding can be fixed-rate, using the same number of bits N_o for each frame, or variable-rate, in which different bit rates are used for different modes. The goal in variable-rate coding is to use only the amount of bits needed to encode the codec parameters to a level adequate to obtain the target quality. As a result, the same target voice quality as that of a fixed-rate, higher-rate coder can be obtained at a significant lower average-rate using variable-bit-rate (VBR) techniques. Conventional VBR speech coders are designed with modes having different bit-rates. An exemplary variable rate speech coder is described in U.S. Patent No. 5,414,796, assigned to the assignee of the present invention and

previously fully incorporated herein by reference. The codec described in the aforesaid patent has the following four rates: (1) full rate (FR); (2) half rate (HR); (3) quarter rate (QR); and (4) eighth rate (ER). For the foregoing rates, each frame of speech is encoded by 160, eighty, forty, and twenty bits per frame, respectively.

- 5 An external open-loop mode decision is made regarding which mode (FR, HR, QR or ER) to apply to the input speech frame.

There is presently a surge of research interest and strong commercial needs to develop a high-quality speech coder operating at medium to low bit rates (i.e., in the range of 2.4 to 4 kbps and below). The application areas include wireless telephony, satellite communications, Internet telephony, various multimedia and voice-streaming applications, voice mail, and other voice storage systems. The driving forces are the need for high capacity and the demand for robust performance under packet loss situations. Various recent speech coding standardization efforts are another direct driving force propelling research and development of low-rate speech coding algorithms. A low-rate speech coder creates more channels, or users, per allowable application bandwidth, and a low-rate speech coder coupled with an additional layer of suitable channel coding can fit the overall bit-budget of coder specifications and deliver a robust performance under channel error conditions.

20 Conventional speech coders typically use some form of prediction mechanism to encode the current frame. Thus, to encode the current frame, a speech coder exploits and uses the information contained in the last decoded and recreated frame. This works well because there is typically strong correlation, or similarity, between successive frames. Thus, a frame or short segment of speech, $S_{cur}(n)$, where $n=1,2,\dots,N$, having N samples can be encoded by a predictive method to form the encoded frame, $S_{cur_quantized}(n)$, according to the following equation:

$$\begin{aligned} S_{cur_quantized}(n) &= S_{cur_predicted}(n) + E_{cur_quantized}(n) \\ 30 \quad &= S_{prev_quantized}(n) * P(n) + E_{cur_quantized}(n), \end{aligned}$$

where "*" represents a convolution operation, $P(n)$ is a conventional prediction filter that produces an approximation of current frame from past quantized frame, $S_{prev_quantized}(n)$ (e.g., $S_{cur_predicted}(n) = S_{prev_quantized}(n) * P(n)$), and $E_{cur_quantized}(n)$ is the quantized version of the prediction error $E_{cur}(n)$ of the current frame. The prediction error is defined as $E_{cur}(n) = S_{cur}(n) - S_{cur_predicted}(n)$.

The performance of the prediction scheme is often measured by a signal-to-noise ratio (SNR) or a perceptual SNR (PSNR), typically defined as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\sum_{n=1}^N W(n) * S_{\text{cur}}(n) * S_{\text{cur}}(n)}{\sum_{n=1}^N W(n) * N_{\text{cur}}(n) * N_{\text{cur}}(n)} \right)$$

5

where $W(n)$, for $n=1,2,\dots,N$, is a perceptual weight factor and $N_{\text{cur}}(n)$ is the error of the overall coding process. The error of the overall coding process is defined as $N_{\text{cur}}(n) = S_{\text{cur}}(n) - S_{\text{cur_quantized}}(n)$. For ordinary SNR, $W(n)$ is set equal to 1 for all $n=1,2,\dots,N$.

10 If the error N_{cur} decreases, the performance of the prediction-based speech coding scheme, or the SNR, will increase. It is therefore advantageous to minimize the error N_{cur} . The equation

$$\begin{aligned} N_{\text{cur}}(n) &= S_{\text{cur}}(n) - S_{\text{cur_quantized}}(n) = [S_{\text{cur}}(n) - S_{\text{cur_predicted}}(n)] + [E_{\text{cur_quantized}}(n) - \\ 15 \quad E_{\text{cur}}(n)] \\ &= \text{Prediction-Error} + \text{Error in-the-Quantization-of-Prediction-Error-Signal} \end{aligned}$$

indicates that the overall error N_{cur} depends on how well the prediction is performed, and how well the prediction error is quantized.

20 The prediction filter information is necessarily sent to the decoder as a certain number of bits, N_p . The remaining available bits, $N_o - N_p$, can be used to encode the prediction error signal E_{cur} . If the prediction from the quantized past frame, $S_{\text{prev_quantized}}$, generates an excellent predicted representation $S_{\text{cur_predicted}}$ of the current frame S_{cur} , the prediction error E_{cur} will be small, having a low dynamic range. Hence, it will be relatively easy to encode the prediction error E_{cur} with a
25 small number of bits.

For high-bit-rate predictive speech coders such as, e.g., the QCELP[®] 13k vocoder manufactured by QUALCOMM INCORPORATED, the total number of bits per frame, N_o , is high. The QCELP[®], for example, supports 260 bits per 20-ms
30 frame. Therefore, even after allocating a number of bits, N_p , to quantize the prediction filter parameter, there are enough remaining bits, $N_o - N_p$, to accurately encode the prediction error. However, at low bit rates (e.g., 4 kbps and below), the total amount of bits available (i.e., eighty or less per frame) is not large enough to

accurately encode both the prediction filter parameters and the prediction error signal. Consequently, the overall coding error N_{cur} grows large, resulting in poor performance and producing a quantized version $S_{cur_quantized}$ of the current frame that could be quite different from the original frame S_{cur} . As the encoding of the
5 next frame depends upon how well the current frame is encoded, the poor performance can degrade the performance of prediction of future frames as well. Thus, there is a need for a variable-rate, multimode, predictive coder that is capable of producing high-voice-quality at low bit rates.

10 SUMMARY OF THE INVENTION

The present invention is directed to a variable-rate, multimode, predictive coder that is capable of producing high-voice-quality at low bit rates. Accordingly, in one aspect of the invention, a speech coder advantageously includes a codec
15 configured to operate in at least one of a plurality of coding modes; and a closed-loop mode decision module coupled to the codec and configured to apply a first coding mode from the plurality of coding modes to an input speech frame, the first coding mode having a first bit rate that is lower than the bit rate of any other coding mode of the plurality of coding modes, the closed-loop mode decision
20 module being further configured to obtain a performance measure of the codec, compare the performance measure with a threshold value, and, if the performance measure does not exceed the threshold value, reject the first coding mode in favor of a second coding mode having a second bit rate that is greater than the first bit rate.

25 In another aspect of the invention, a method of coding speech frames advantageously includes the steps of selecting a first coding mode to apply to a speech frame, the first coding mode having a first bit rate; obtaining a coding performance measure; comparing the coding performance measure with a threshold value; and rejecting the first coding mode in favor of a second coding
30 mode if the coding performance measure does not exceed the threshold value, the second coding mode having a second bit rate that exceeds the first bit rate.

In another aspect of the invention, a speech coder advantageously includes means for selecting a first coding mode to apply to a speech frame, the first coding mode having a first bit rate; means for obtaining a coding performance measure;
35 means for comparing the coding performance measure with a threshold value; and means for rejecting the first coding mode in favor of a second coding mode if the

coding performance measure does not exceed the threshold value, the second coding mode having a second bit rate that exceeds the first bit rate.

BRIEF DESCRIPTION OF THE DRAWINGS

5

FIG. 1 is a block diagram of a communication channel terminated at each end by speech coders.

FIG. 2 is a block diagram of an encoder.

FIG. 3 is a block diagram of a decoder.

10

FIG. 4 is a flow chart illustrating the steps of a closed-loop, multimode, predictive coding technique for speech frames at low bit rates.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

15

In FIG. 1 a first encoder 10 receives digitized speech samples $s(n)$ and encodes the samples $s(n)$ for transmission on a transmission medium 12, or communication channel 12, to a first decoder 14. The decoder 14 decodes the encoded speech samples and synthesizes an output speech signal $s_{\text{SYNTH}}(n)$. For transmission in the opposite direction, a second encoder 16 encodes digitized speech samples $s(n)$, which are transmitted on a communication channel 18. A second decoder 20 receives and decodes the encoded speech samples, generating a synthesized output speech signal $s_{\text{SYNTH}}(n)$.

20

The speech samples $s(n)$ represent speech signals that have been digitized and quantized in accordance with any of various methods known in the art including, e.g., pulse code modulation (PCM), companded μ -law, or A-law. As known in the art, the speech samples $s(n)$ are organized into frames of input data wherein each frame comprises a predetermined number of digitized speech samples $s(n)$. In an exemplary embodiment, a sampling rate of 8 kHz is employed, with each 20 ms frame comprising 160 samples. In the embodiments described below, the rate of data transmission may advantageously be varied on a frame-to-frame basis from 8 kbps (full rate) to 4 kbps (half rate) to 2 kbps (quarter rate) to 1 kbps (eighth rate). Varying the data transmission rate is advantageous because lower bit rates may be selectively employed for frames containing relatively less speech information. As understood by those skilled in the art, other sampling rates, frame sizes, and data transmission rates may be used.

30

35

The first encoder 10 and the second decoder 20 together comprise a first speech coder, or speech codec. Similarly, the second encoder 16 and the first decoder 14 together comprise a second speech coder. It is understood by those of skill in the art that speech coders may be implemented with a digital signal processor (DSP), an application-specific integrated circuit (ASIC), discrete gate logic, firmware, or any conventional programmable software module and a microprocessor. The software module could reside in RAM memory, flash memory, registers, or any other form of writable storage medium known in the art. Alternatively, any conventional processor, controller, or state machine could be substituted for the microprocessor. Exemplary ASICs designed specifically for speech coding are described in U.S. Patent No. 5,727,123, assigned to the assignee of the present invention and fully incorporated herein by reference, and U.S. Application Serial No. 08/197,417, entitled VOCODER ASIC, filed February 16, 1994, assigned to the assignee of the present invention, and fully incorporated herein by reference.

In FIG. 2 an encoder 100 that may be used in a speech coder includes a mode decision module 102, a pitch estimation module 104, an LP analysis module 106, an LP analysis filter 108, an LP quantization module 110, and a residue quantization module 112. Input speech frames $s(n)$ are provided to the mode decision module 102, the pitch estimation module 104, the LP analysis module 106, and the LP analysis filter 108. The mode decision module 102 produces a mode index I_M and a mode M based upon the periodicity of each input speech frame $s(n)$. Various methods of classifying speech frames according to periodicity are described in U.S. Application Serial No. 08/815,354, entitled METHOD AND APPARATUS FOR PERFORMING REDUCED RATE VARIABLE RATE VOCODING, filed March 11, 1997, assigned to the assignee of the present invention, and fully incorporated herein by reference. Such methods are also incorporated into the Telecommunication Industry Association Industry Interim Standards TIA/EIA IS-127 and TIA/EIA IS-733.

The pitch estimation module 104 produces a pitch index I_p and a lag value P_0 based upon each input speech frame $s(n)$. The LP analysis module 106 performs linear predictive analysis on each input speech frame $s(n)$ to generate an LP parameter a . The LP parameter a is provided to the LP quantization module 110. The LP quantization module 110 also receives the mode M . The LP quantization module 110 produces an LP index I_L and a quantized LP parameter \hat{a} . The LP analysis filter 108 receives the quantized LP parameter \hat{a} in addition to the input

speech frame $s(n)$. The LP analysis filter 108 generates an LP residue signal $R[n]$, which represents the error between the input speech frames $s(n)$ and the quantized linear predicted parameters \hat{s} . The LP residue $R[n]$, the mode M , and the quantized LP parameter \hat{a} are provided to the residue quantization module 112.

5 Based upon these values, the residue quantization module 112 produces a residue index I_R and a quantized residue signal \hat{r} .

In FIG. 3 a decoder 200 that may be used in a speech coder includes an LP parameter decoding module 202, a residue decoding module 204, a mode decoding module 206, and an LP synthesis filter 208. The mode decoding module 206

10 receives and decodes a mode index I_M , generating therefrom a mode M . The LP parameter decoding module 202 receives the mode M and an LP index I_{LP} . The LP parameter decoding module 202 decodes the received values to produce a quantized LP parameter \hat{a} . The residue decoding module 204 receives a residue index I_R , a pitch index I_P , and the mode index I_M . The residue decoding module 204

15 decodes the received values to generate a quantized residue signal \hat{r} . The quantized residue signal $\hat{r}[n]$ and the quantized LP parameter \hat{a} are provided to the LP synthesis filter 208, which synthesizes a decoded output speech signal $\hat{s}[n]$ therefrom.

Operation and implementation of the various modules of the encoder 100 of FIG. 2 and the decoder of FIG. 3 are known in the art, and are described in detail in L.B. Rabiner & R.W. Schafer *Digital Processing of Speech Signals* 396-453 (1978), which is fully incorporated herein by reference. An exemplary encoder and an exemplary decoder are described in U.S. Patent No. 5,414,796, previously fully incorporated herein by reference.

25 In one embodiment a multimode coder first uses an open-loop decision mode, relying on parameters extracted out of the current frame to classify the current frame as background-noise/silence (N), unvoiced speech (UV), or voiced speech (V). Various speech classification methods used for rate determination are known in the art, including methods described in the aforementioned U.S. Patent

30 No. 5,414,796, previously fully incorporated herein by reference. N-type frames are coded with an eighth-rate mode, and UV-type frames are coded with a quarter-rate mode.

For V-type frames (i.e., voiced speech frames), either a higher-rate ($N_0=N_1$ bits per frame) mode such as full rate, or a lower-rate ($N_0=N_2$ bits per frame, where $N_2 < N_1$) mode such as half rate, is used. The full-rate mode may

35 advantageously be a prediction-based coding scheme with adequate bits to

accurately encode various types of voiced speech, delivering a perceptual signal-to-noise ratio (PSNR) well above the target PSNR (a predefined or variable threshold value). The half-rate mode is advantageously a prediction-based coding scheme designed to encode frames a high degree of correlation with the previous frame (i.e., frames that are quite similar to the previous frame). Thus, the number of bits available in the half-rate mode, N2 bits per frame, is adequate to encode the prediction parameters for frames with high correlation, as well as the prediction error, which is relatively small due to the high correlation between successive frames. Such frames are typically encountered in steady voiced speech segments, which are therefore amenable to half-rate coding. Additionally, the performance of prediction-based coding schemes also depends on how accurately the previous frame is quantized. Hence, a closed-loop mode selection process is employed after the open-loop mode to ensure that the coding performance exceeds the predefined (or variable) target PSNR value. As those of skill in the art would understand, the open-loop mode need not necessarily be applied at all.

The flow chart of FIG. 4 illustrates a closed-loop, multimode, predictive coding technique for speech frames at low bit rates, in accordance with one embodiment. In step 300 a frame number counter is set equal to 1. The algorithm then proceeds to step 302, starting the coding process. The algorithm then proceeds to step 304. In step 304 the algorithm checks the current frame and the previous quantized frame. The algorithm then proceeds to step 306. In step 306 the algorithm determines whether the current frame should be classified as silence or background noise. This determination is made in accordance with various conventional techniques for measuring frame energy, such as, e.g., calculating the sum-of-squares. If the frame is classified as silence or background noise, the algorithm proceeds to step 308. In step 308 the algorithm applies an eighth-rate coding mode to the frame. The algorithm then proceeds to step 310. If, on the other hand, in step 306 the frame is not classified as background noise or silence, the algorithm proceeds to step 312.

In step 312 the algorithm determines whether the current frame should be classified as unvoiced speech. This determination is made in accordance with various known methods of periodicity determination, such as, e.g., the use of zero crossings and normalized autocorrelation functions (NACFs). These techniques are described in the aforementioned U.S. Application Serial No. 08/815,354, previously fully incorporated herein by reference. If the frame is classified as unvoiced speech, the algorithm proceeds to step 314. In step 314 a quarter-rate

coding mode is applied to the frame. The algorithm then proceeds to step 310. If, on the other hand, in step 312 the frame is not classified as unvoiced speech, the algorithm proceeds to step 316, considering the frame to contain voiced speech. In step 316 the algorithm goes to a half-rate prediction-based coding mode. The
5 algorithm then proceeds to step 318. In step 318 the PSNR is computed. The algorithm then proceeds to step 320.

In step 320 the algorithm determines whether the computed PSNR is greater than a predefined threshold, or target, PSNR value. As an alternative, the threshold, or target, PSNR value may be a function of average bit rate. For
10 example, the average bit rate is calculated periodically and fed back to the algorithm, which adjusts the target threshold value accordingly. Further, it should be understood that any conventional measure of performance may be substituted for PSNR. If the computed PSNR exceeds the target PSNR, the algorithm proceeds to step 322. In step 322 a half-rate coding mode is applied to the frame. The
15 algorithm then proceeds to step 310. If, on the other hand, in step 320 the computed PSNR does not exceed the target PSNR, the algorithm proceeds to step 324. In step 324 the algorithm applies a full-rate coding mode to the frame. The algorithm then proceeds to step 310.

In step 310 the frame number counter is incremented by 1. The algorithm
20 then proceeds to step 326. In step 326 the algorithm determines whether the frame number counter value is greater than or equal to the total number of frames that must be processed (i.e., whether there are any remaining frames to process). If the frame number counter value is less than the total number of frames to be processed, the algorithm returns to step 302, beginning the coding process for the
25 next frame. If, on the other hand, the frame number counter value is greater than or equal to the total number of frames to be processed, the algorithm proceeds to step 328, ending the coding process.

In alternate embodiments the full-rate coding mode described above with respect to FIG. 4 could be a higher-bit-rate predictive mechanism (i.e., any bit rate
30 that is greater than half-rate). In one embodiment a higher-bit-rate, direct coding mechanism is substituted for the full-rate, predictive coding mode. The direct coding mode encodes the current speech frame or residue without using any information from the previous frame.

The use of a direct encoding method is appropriate for speech segments for
35 which there is no similarity between the current frame and the previous frame. An example is during the onset of a voice segment. Another example is unvoiced-to-

voiced segment transitions. A direct encoding method is also useful in the middle of voiced segments when the cumulative effect of prediction-based encoding has degraded the past quantized frame so as to be too far out of sync with the corresponding original speech frame. In this case predictive coding will fail, even
5 at much higher bit rates, due to the lack of similarity between the past quantized frame and the past original frame. In such a case, a fresh capture of the current frame with a direct encoding method will not only enhance the preservation of the current frame, but will also facilitate future prediction-based encoding of the next and later frames because the prediction mechanism will be aided by a more
10 accurate memory.

Those of skill would understand that while the embodiments described above contemplate four bit rates, any reasonable number of bit rates could be substituted for four. Those of skill would further appreciate that the embodiments described herein could be extended to analysis over a number of frames that is
15 greater than one, at the expense of additional processing time or capability.

In one embodiment two modes may be employed, with bit rates R1 and R2. The R1 coding method is a higher-rate, direct coding method. The R2 coding method is a lower-rate, predictive coding method. A closed-loop decision is performed such that the R2 coding method is tried first, the performance is
20 checked by comparing with a performance measure, and the algorithm switches to the R1 coding method if the performance for the R2 coding mode is insufficient. In an alternate embodiment, the higher-rate, R1 coding mode is tried first, the performance is checked by comparing with a performance measure, and, if the performance is satisfactory, the lower-rate, R2 coding mode is tried. The
25 performance check is then performed for the R2 coding mode, and if the R2 coding mode performance is unsatisfactory, the R1 coding mode is applied to the frame.

In another embodiment multiple coding modes having bit rates R1, R2, ..., RN-1, RN (where $R1 > R2 > \dots > RN-1 > RN$) are employed. A closed-loop decision is performed such that the lowest rate, RN, is tried first. If the RN coding
30 mode performs adequately, the RN coding mode is retained for the frame. Otherwise, the next, higher-rate coding mode, RN-1, is applied. The process is reiterated until either a coding mode performs adequately or the highest-rate mode, R1, is retained. In an alternate embodiment, the highest rate, R1, is tried first. If the R1 mode performs adequately, the next, lower-rate coding mode, R2, is
35 tried. The process is continued until a given coding mode does not perform

adequately (at which time the last coding mode to perform adequately is applied), or until the lowest-rate coding mode, R_N , performs satisfactorily and is applied.

In another embodiment multiple coding modes having bit rates $R_1, R_2, \dots, R_{m-1}, R_m, R_{m+1}, \dots, R_N$ are employed. The bit rates have the following relative magnitudes: $R_1 > R_2 > \dots > R_{m-1} > R_m > R_{m+1} > \dots > R_N$. A closed-loop mode decision works in conjunction with an open-loop mode decision. The open-loop mode decision, based upon parameters such as frame energy or frame periodicity, tells the coder to apply a mode with a bit rate of R_m , at which point the closed-loop mode decision takes over. The closed-loop mode decision applies the R_m coding mode, tests performance, and maintains the R_m coding mode if performance is satisfactory. Otherwise, the closed-loop mode decision tries the next, higher-rate coding mode, R_{m-1} . The process is reiterated until either a coding mode performs adequately or the highest-rate mode, R_1 , is retained. Alternatively, the closed-loop mode decision applies the R_m coding mode, tests performance, and maintains the R_m coding mode if performance is satisfactory. Otherwise, the closed-loop mode decision tries the next, lower-rate coding mode, R_{m+1} . The process is reiterated until either a coding mode performs inadequately (at which time the last coding mode to perform adequately is applied), or the lowest-rate mode, R_N , is retained.

In another embodiment multiple coding modes having bit rates R_1, R_2, \dots, R_N (where $R_1 > R_2 > \dots > R_N$) are employed. All of the coding modes are applied in parallel to the input speech frame, and the performances of the coding modes are compared with a set of N threshold performance measures. The coding mode that appears to produce the most accurate result is selected.

In another embodiment multiple coding modes having bit rates R_1, R_2, \dots, R_N (where $R_1 > R_2 > \dots > R_N$) are employed. All of the coding modes are applied in parallel to the input speech frame, and the performances of the coding modes are compared with a set of N threshold performance measures. If several coding modes exceed the performance threshold target, the coding mode having the lowest bit rate (and also performing above the performance threshold) is selected.

In another embodiment multiple coding modes having bit rates $R_1, R_2, \dots, \text{Quarter Rate}, \dots, \text{Half Rate}, \dots, R_N$ (where R_1 is Full Rate and R_N is Eighth Rate) are employed. A closed-loop mode decision works in conjunction with an open-loop mode decision. The open-loop mode decision, based upon parameters such as frame energy or frame periodicity, tells the coder to apply the full-rate coding mode to unvoiced-to-voiced transition frames, voiced-to-voiced transition

frames, nonstationary voiced segments, and nonstationary unvoiced segments. Also based upon frame parameters, the open-loop mode decision tells the coder to apply the half-rate coding mode to steady-voiced segments that exhibit a significant degree of similarity from frame to frame. Also based upon frame parameters, the open-loop mode decision tells the coder to apply the quarter-rate coding mode to steady unvoiced segments. Also based upon frame parameters, the open-loop mode decision tells the coder to apply the eighth-rate coding mode to background noise and other nonspeech signals such as silence. Once the open-loop mode decision has selected a coding mode for application to the frame, the closed-loop mode decision takes over. The closed-loop mode decision applies the coding mode selected by the open-loop mode decision, tests performance, and maintains the selected coding mode if performance is satisfactory. Otherwise, the closed-loop mode decision tries the next, higher-rate coding mode. The process is reiterated until either a coding mode performs adequately or the full-rate mode is retained. Alternatively, the closed-loop mode decision applies the coding mode selected by the open-loop mode decision, tests performance, and maintains the selected coding mode if performance is satisfactory. Otherwise, the closed-loop mode decision tries the next, lower-rate coding mode. The process is reiterated until either a coding mode performs inadequately (at which time the last coding mode to perform adequately is applied), or the lowest-rate mode is retained.

In another embodiment a multimode coder includes a first set of N modes, M_i , and the first set of modes has respective bit rates R_i , where $i=1,2,\dots,N$. The coder also has a second set of N modes, MCC_i , and the second set of modes has respective bit rates RCC_i , where $i=1,2,\dots,N$. The MCC_i and M_i coding modes each use the same source-coding mode (i.e., the same encoder and decoder). However, the MCC_i coding mode includes an additional layer of channel protection, in which $(RCC_i - R_i)$ bits are used for robust protection of the parameters of the M_i coding mode under the worst possible channel condition of the communication system. Hence, the performance, or voice quality, delivered by the M_i coding mode under channel-error-free conditions is similar to the performance, or voice quality, delivered by the MCC_i coding mode under the worst possible channel error condition. The $(RCC_i - R_i)$ channel coding bits serve to provide adequate protection under the assumed, or target, worst channel condition. The assumed worst channel condition may advantageously be, e.g., a predefined percentage of frame error rate (FER).

In this particular embodiment, a closed-loop mode decision advantageously accounts for both channel variation and source variation to deliver a guaranteed quality of service. For example, a source-controlled, closed-loop mode decision such as described above is applied first. The closed-loop mode decision tells the
5 coder to use the M_i coding mode. An external, network-control indicator SW , which is a signal provided by the communication network to the speech encoder, indicates whether the communication channel is in good condition (e.g., if $SW=1$, the channel is error-free) in bad condition (e.g., if $SW=0$, the channel is erroneous). If the channel is in good condition, the coding mode M_i , having bit rate R_i , is used.
10 If, on the other hand, the channel is in bad condition, the coding mode MCC_i , having bit rate RCC_i , is used.

Those skilled in the art would appreciate that the number of network conditions need not be restricted to two. Thus, in one embodiment, a multimode
15 coder is designed to account for $j=1,2,\dots,M$ different possible network conditions by providing M different modes $MCC_{i,j}$ having rates $RCC_{i,j}$, where $j=1,2,\dots,M$, for each original source-controlled coding mode M_i . Such a scheme allows for varied amounts of channel coding because $(RCC_{i,j}-RCC_i)$ represents the minimum number of bits needed to add channel error protection to the channel coding layer so that the channel error protection will be adequate for the worst-case scenario in
20 the j -th channel error condition. The source-controlled, closed-loop mode decision then determines which coding mode M_i to apply first, and, based on the value of $SW=j$ (where $j=1,2,\dots,M$), selects the coding mode $MCC_{i,j}$. Such a closed-loop, combined-network-and-source-controlled codec delivers guaranteed quality of service across various channel conditions while also delivering a low average bit
25 rate.

Preferred embodiments of the present invention have thus been shown and described. It would be apparent to one of ordinary skill in the art, however, that numerous alterations may be made to the embodiments herein disclosed without departing from the spirit or scope of the invention. Therefore, the present
30 invention is not to be limited except in accordance with the following claims.

What is claimed is:

CLAIMS

1. A speech coder, comprising:
 - 2 a codec configured to operate in at least one of a plurality of coding modes; and
 - 4 a closed-loop mode decision module coupled to the codec and configured to apply a first coding mode from the plurality of coding modes to an
 - 6 input speech frame, the first coding mode having a first bit rate that is lower than the bit rate of any other coding mode of the plurality of coding modes, the closed-
 - 8 loop mode decision module being further configured to obtain a performance measure of the codec, compare the performance measure with a threshold value, and, if the performance measure does not exceed the threshold value, reject the
 - 10 first coding mode in favor of a second coding mode having a second bit rate that is
 - 12 greater than the first bit rate.
2. The speech coder of claim 1, wherein the closed-loop mode decision
- 2 module is configured to continue a process of selecting and, based on performance, rejecting coding modes chosen successively in order of increasing bit rate.
3. The speech coder of claim 1, wherein the performance based measure
- 2 is obtained by comparing a resultant synthetic speech frame with the input speech frame.
4. The speech coder of claim 1, wherein the first coding mode is a
- 2 prediction-based coding mode and the second coding mode is a direct coding mode.
5. The speech coder of claim 1, further comprising an open-loop mode
- 2 decision module coupled to the codec and configured to select one of the plurality of coding modes for application to the input speech frame before the closed-loop
- 4 mode decision module applies a coding mode, wherein the closed-loop mode decision module is configured to first apply the coding mode selected by the open-
- 6 loop mode decision module.

6. The speech coder of claim 2, further comprising an open-loop mode decision module coupled to the codec and configured to select one of the plurality of coding modes for application to the input speech frame before the closed-loop mode decision module applies a coding mode, wherein the closed-loop mode decision module is configured to first apply the coding mode selected by the open-loop mode decision module.

7. The speech coder of claim 1, wherein the threshold value is a predefined quantity.

8. The speech coder of claim 1, wherein the threshold value is a function of average bit rate.

9. A method of coding speech frames, comprising the steps of:
selecting a first coding mode to apply to a speech frame, the first coding mode having a first bit rate;
obtaining a coding performance measure;
comparing the coding performance measure with a threshold value;
and
rejecting the first coding mode in favor of a second coding mode if the coding performance measure does not exceed the threshold value, the second coding mode having a second bit rate that exceeds the first bit rate.

10. The method of claim 9, further comprising the step of repeating the obtaining, comparing, and rejecting steps in successive order until the coding performance measure exceeds the threshold value.

11. The method of claim 9, wherein the obtaining step comprises comparing a resultant synthetic speech frame with the speech frame.

12. The method of claim 9, wherein the first coding mode is a prediction-based coding mode and the second coding mode is a direct coding mode.

13. The method of claim 9, wherein the selecting step comprises selecting a first coding mode based upon parameters of the speech frame.

14. The method of claim 10, wherein the selecting step comprises
2 selecting a first coding mode based upon parameters of the speech frame.

15. The method of claim 9, wherein the comparing step comprises
2 comparing the coding performance measure with a predefined threshold value.

16. The method of claim 9, wherein the comparing step comprises
2 comparing the coding performance measure with a threshold value that is a
function of average bit rate.

17. A speech coder, comprising:
2 means for selecting a first coding mode to apply to a speech frame,
the first coding mode having a first bit rate;
4 means for obtaining a coding performance measure;
means for comparing the coding performance measure with a
6 threshold value; and
means for rejecting the first coding mode in favor of a second coding
8 mode if the coding performance measure does not exceed the threshold value, the
second coding mode having a second bit rate that exceeds the first bit rate.

18. The speech coder of claim 17, further comprising means for
2 continuing to obtain the performance measure, compare the performance measure
with the threshold value, and reject coding modes in favor of other coding modes
4 having greater bit rates until the coding performance measure exceeds the
threshold value.

19. The speech coder of claim 17, wherein the means for obtaining
2 comprises means for comparing a resultant synthetic speech frame with the speech
frame.

20. The speech coder of claim 17, wherein the first coding mode is a
2 prediction-based coding mode and the second coding mode is a direct coding
mode.

21. The speech coder of claim 17, wherein the means for selecting
2 comprises means for selecting a first coding mode based upon parameters of the
speech frame.

22. The speech coder of claim 18, wherein the means for selecting
2 comprises means for selecting a first coding mode based upon parameters of the
speech frame.

23. The speech coder of claim 17, wherein the threshold value is a
2 predefined quantity.

24. The speech coder of claim 17, wherein the threshold value is a
2 function of average bit rate.

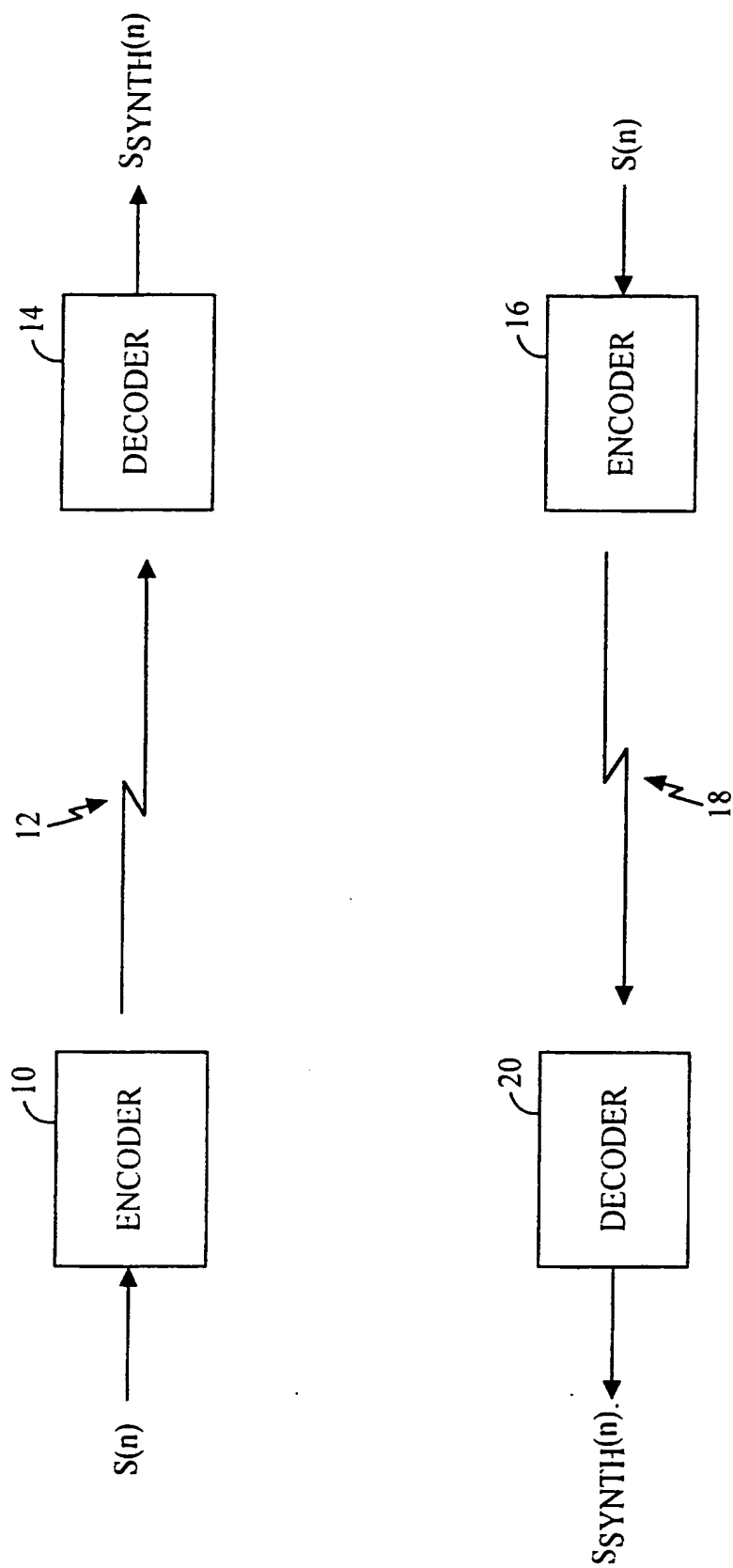


FIG. 1

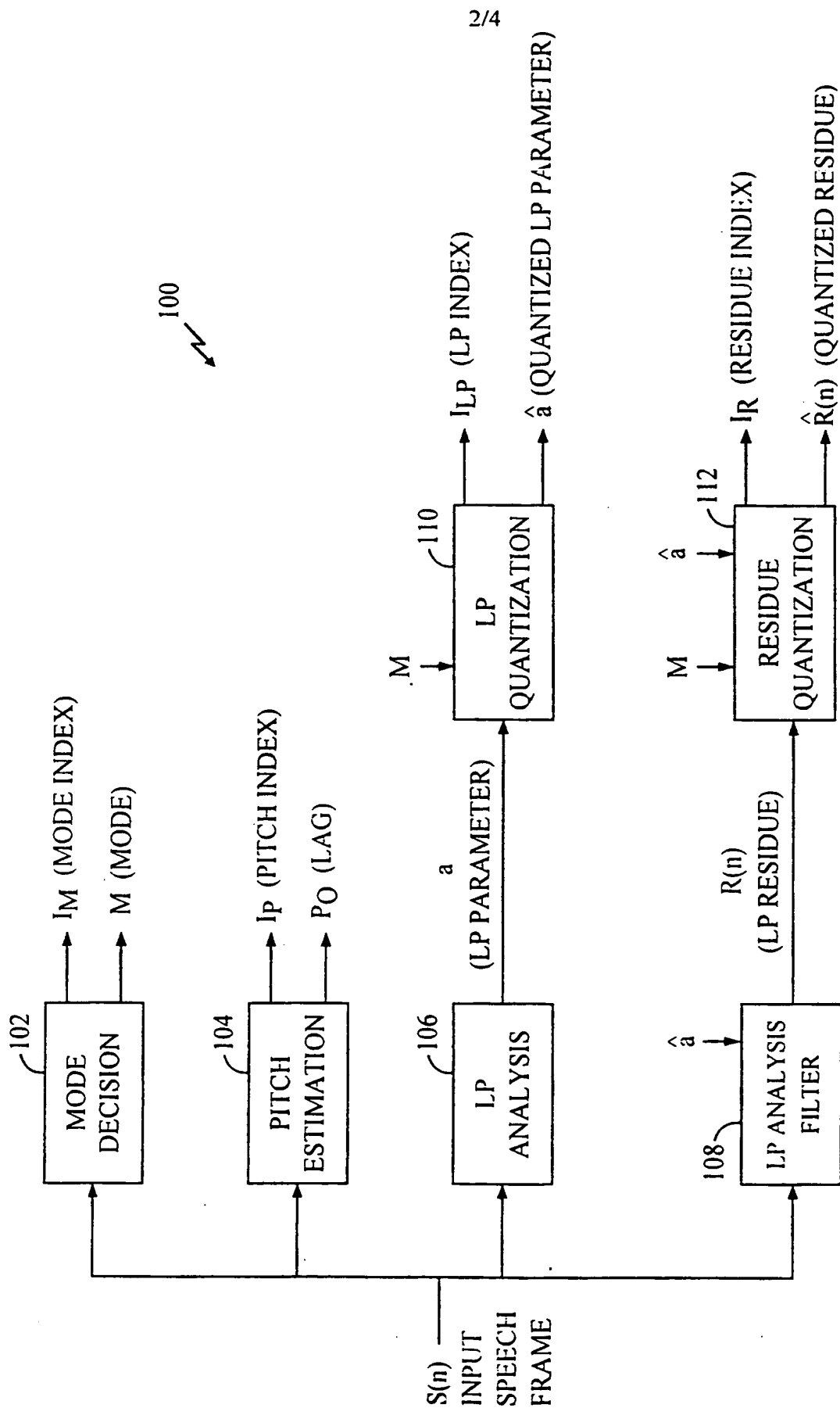


FIG. 2

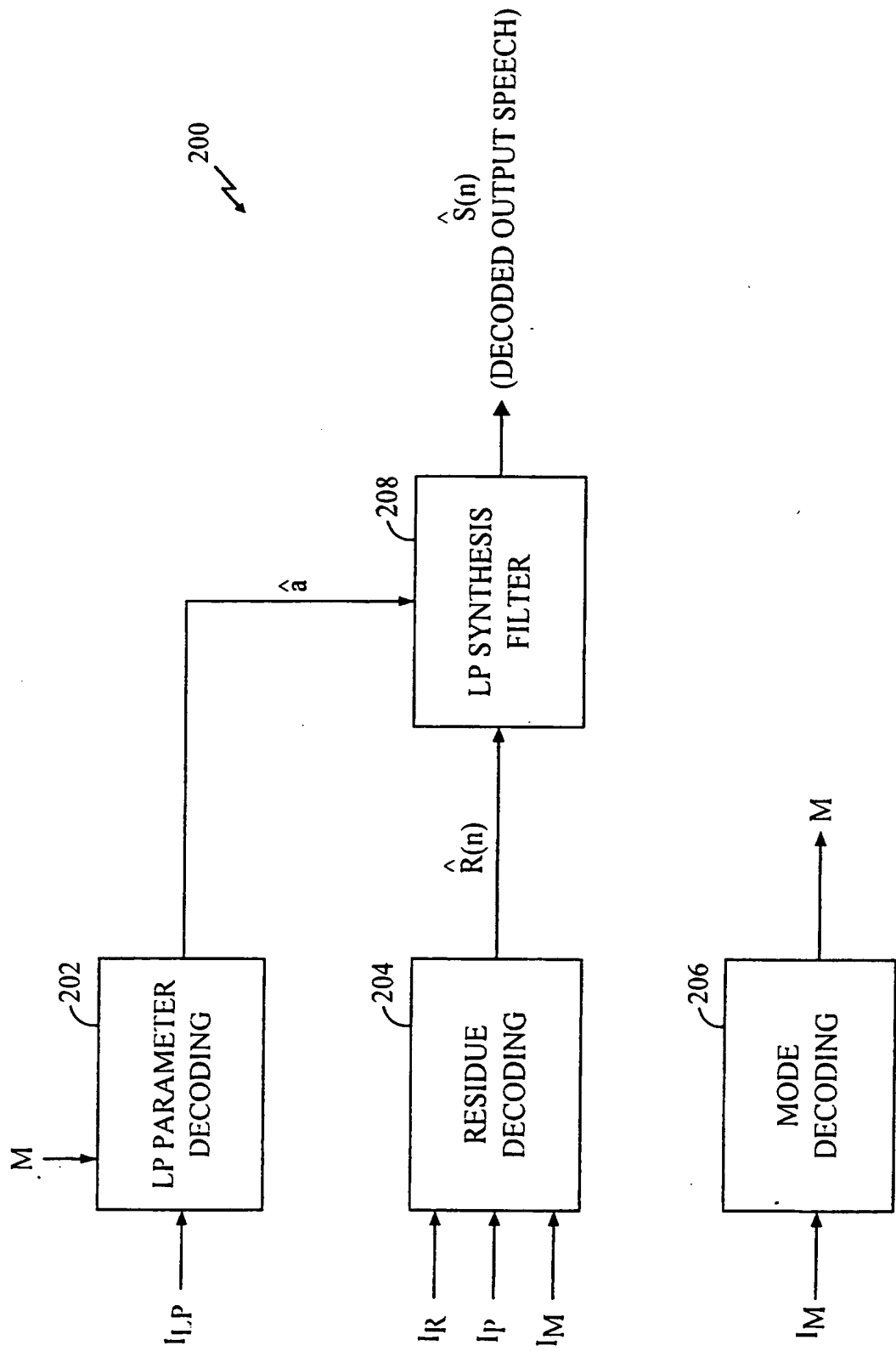


FIG. 3

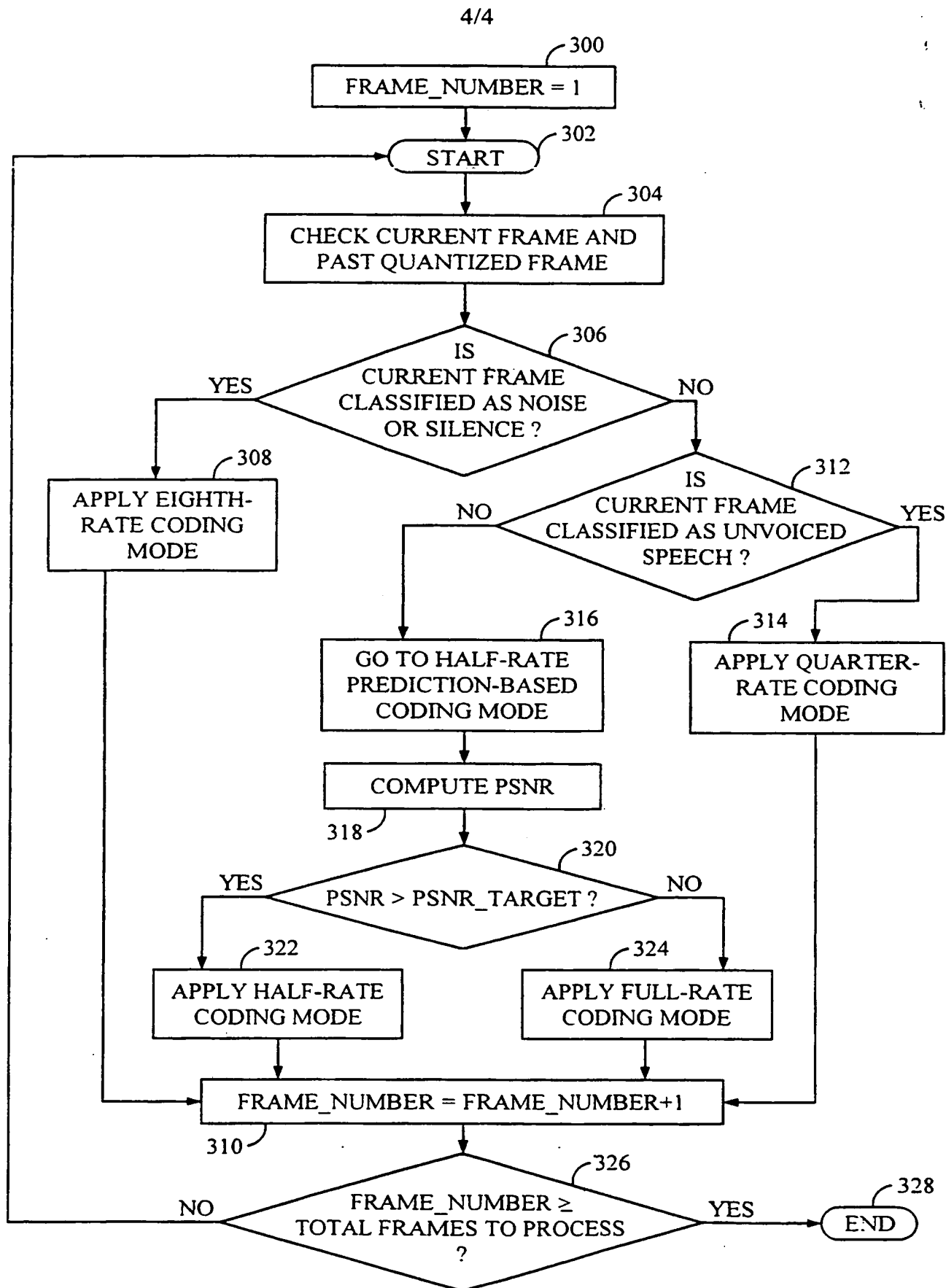


FIG. 4.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/26850

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L19/14

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CELLARIO L ET AL: "CELP CODING AT VARIABLE RATE" EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS AND RELATED TECHNOLOGIES, IT, AEI, MILANO, vol. 5, no. 5, 1 September 1994 (1994-09-01), pages 69-79, XP000470681 ISSN: 1120-3862 paragraph '0004!; figure 2 paragraph '04.3!; figure 6	1-3, 5-7, 9-11, 13-15, 17-19, 21-23
X	US 5 761 634 A (STEWART KENNETH A ET AL) 2 June 1998 (1998-06-02) abstract figure 4	1-3, 9-11, 17-19

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

29 February 2000

Date of mailing of the international search report

14/03/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3018

Authorized officer

Krembel, L

INTERNATIONAL SEARCH REPORT

Inter. Appl. No.

PCT/US 99/26850

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>DAS A ET AL: "Multimode variable bit rate speech coding: an efficient paradigm for high-quality low-rate representation of speech signal"</p> <p>1999 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. ICASSP99 (CAT. NO.99CH36258), 1999 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. ICASSP99, PHOENIX, AZ, USA, 15-19 , pages 2307-2310 vol.4, XP002131859 1999, Piscataway, NJ, USA, IEEE, USA ISBN: 0-7803-5041-3</p> <p>paragraph '02.3!; figure 2</p>	1-24
A	<p>EP 0 417 739 A (FUJITSU LTD)</p> <p>20 March 1991 (1991-03-20)</p> <p>abstract</p> <p>figure 2</p>	1,9,17
A	<p>KLEIDER J E ET AL: "AN ADAPTIVE-RATE DIGITAL COMMUNICATION SYSTEM FOR SPEECH"</p> <p>IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP),US,NEW YORK, IEEE,1997, pages 1695-1698, XP000734984 ISBN: 0-8186-7920-4</p> <p>paragraph '02.E!</p> <p>figure 1</p>	1,9,17

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/US 99/26850

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5761634	A	02-06-1998	CA 2157024 A	24-08-1995
			CN 1121375 A	24-04-1996
			EP 0699334 A	06-03-1996
			FI 954622 A	29-09-1995
			JP 8509348 T	01-10-1996
			WO 9522818 A	24-08-1995
EP 0417739	A	20-03-1991	JP 3098318 A	23-04-1991
			CA 2024742 A, C	12-03-1991
			DE 69020269 D	27-07-1995
			DE 69020269 T	25-01-1996
			US 5224167 A	29-06-1993